AMENDMENTS TO THE CLAIMS:

This listing of claims will replace all prior versions and listings of claims in the application:

Listing of Claims:

1.   (Currently amended) A device for automatically identifying the language of a digital text, comprising:

means for prestoring first character strings, including prefixes, suffixes and infixes, of different lengths from words of a plurality of predetermined languages, that occur frequently anywhere respectively in said words of said plurality of predetermined languages,

means for prestoring second character strings of different lengths, that are atypical anywhere respectively in said words of said predetermined languages,

means for analyzing words extracted from said digital text, thereby constructing for each extracted word a-plurality-of all the character strings contained in said extracted word, including all the prefixes, suffixes and infixes in said extracted word, with overlap and different lengths lying between one character and the number of characters in said extracted word, and

means for comparing each of said plurality-of character strings contained in each said extracted word to said first and-second-prestored character strings and second prestored character strings of said predetermined languages,

2

Application No.: 10/732,809   **Docket No.: 324-163**

means for calculating scores respectively associated with said predetermined languages ~~so that whenever a first character string is found in said extracted word~~, a score associated with one determined language being calculated by adding to said score ~~is increased by~~ a first coefficient whenever a prestored ~~depending on the position of said~~ first character string of said one determined language is found in said extracted word, said first coefficient depending on the position of said found prestored first character string of said one determined language in said extracted word, and, by subtracting from said score a second coefficient whenever a prestored second character string of said one determined language is found in said extracted word, ~~said score is decreased by a respective second coefficient that is associated with said found second character string, said respective~~ second coefficient increasing as the probability of said found prestored second character string in said one determined language decreases, and

means for comparing said scores for said text associated with said predetermined languages in order to determine the highest of said scores, which identifies the language of said text.

2.   (Cancelled)

3.   (Original) The device claimed in claim 1, wherein said first coefficient of a first character string in said extracted word depends on the frequency of said character string in said determined language.

3

4. (Original) The device claimed in claim 1, wherein said first coefficient of a first character string in said extracted word depends on the length of said character string.

5. (Original) The device claimed in claim 1, wherein said first coefficient of a first character string in said extracted word is equal to:

$$PO\ (FR+LON),$$

where

PO is a coefficient depending on the position of said first character string in said extracted word,

FR is a coefficient depending on the frequency of said first character string in a determined language, and

LON is a coefficient depending on the length of said first character string.

6. (Previously presented) The device claimed in claim 1, comprising comparator means for comparing each of said extracted words from said text with frequent words in said determined language and initially listed in storage means so that whenever a frequent word is found in said text, said score for said determined language is increased only by a coefficient depending on the frequency of said extracted word in said determined language.

7. (Previously presented) The device claimed in claim 1, comprising comparator means for comparing each of said extracted words from said text with frequent words in said determined language and initially listed in storage means so that

4

whenever a frequent word is found in said text, said score for said determined language is increased only by a coefficient depending on the length of said frequent word.

8.   (Cancelled)

9.   (Cancelled)

10.   (Currently amended)   The method of claim [[8]]13, wherein said first coefficient of a first character string in said extracted word is equal to:

$$PO (FR+LON),$$

where

PO is a coefficient depending on the position of said first character string in said extracted word,

FR is a coefficient depending on the frequency of said first character string in a determined language, and

LON is a coefficient depending on the length of said first character string.

11.   (cancelled)

12.   (currently amended) A device for automatically identifying the language of a digital text, comprising:

means for prestoring first character strings that occur frequently anywhere respectively in words of a plurality of predetermined languages and characterize said predetermined languages,

means for prestoring second character strings that are atypical anywhere respectively in words of said predetermined languages,

5

means for analyzing words extracted from said digital text, thereby constructing for each extracted word all character strings contained in said extracted word and having lengths lying between one character and the number of characters in said extracted word,

means for comparing character strings contained in extracted words to prestored character strings in order to determine scores associated with said predetermined languages,

means for individually comparing each of all character strings contained in each said extracted word to said first prestored character strings and said second prestored character strings of each determined language so that whenever a prestored first character string is found in said extracted word, a score associated with said each determined language is increased by a first coefficient depending on the position of said first character string found in said extracted word, and, whenever a prestored second character string is found in said extracted word, said score is decreased by a respective second coefficient that is associated with said found second character string, said respective second coefficient increasing as the probability of said found second character string in said each determined language decreases, and

means for comparing said scores for said text associated with said predetermined languages in order to determine the highest of said scores, which identifies the language of said text, wherein said first coefficient of a first character string in said extracted word is equal to:

$$PO \ (FR+LON),$$

6

where

PO is a coefficient depending on the position of said first character string in said extracted word,

FR is a coefficient depending on the frequency of said first character string in a determined language, and

LON is a coefficient depending on the length of said first character string.

13. (New) A method of automatically identifying the language of a digital text, the method being performed with a computer arrangement including a storing arrangement and a processor arrangement,

the storing arrangement prestoring (a) first character strings, including prefixes, suffixes and infixes, of different lengths from words of a plurality of predetermined languages, that occur frequently anywhere respectively in said words of said plurality of predetermined languages, and (b) second character strings of different lengths, that are atypical anywhere respectively in said words of said predetermined languages,

the method comprising:

in the processor arrangement: (a) analyzing words extracted from said digital text, thereby constructing for each extracted word all the character strings contained in said extracted word, including all the prefixes, suffixes and infixes in said extracted word, with overlap and different lengths lying between one character and the number of characters in said extracted word, (b) comparing each of said character strings contained in each said extracted word with said first prestored character strings and second prestored character strings of said predetermined languages, (c) calculating scores respectively associated with said predetermined languages, the calculation of a score associated with one determined language being performed (i) by adding to said score a first coefficient whenever a prestored first character string of said one determined language is found in said extracted word, said first coefficient depending on the position of said found prestored first character string of said one determined

7

language in said extracted word, and (ii) subtracting from said score a second coefficient whenever a prestored second character string of said one determined language is found in said extracted word, said second coefficient thereby increasing as the probability of said found prestored second character string in said one determined language decreases, and (d) comparing said scores for said text associated with said predetermined languages in order to determine the highest of said scores, the highest of the scores identifying the language of said text.

8